杜仲基因密码子使用模式分析

王艳[1]，赵懿琛[1*]，赵德刚[1,2]

（1. 贵州大学 生命科学学院，茶学院，山地植物资源保护与种质创新省部共建教育部重点实验室，贵阳 550000；2. 贵州省农业科学研究院，贵阳 550000）

摘要：为了解杜仲基因密码子使用模式，以杜仲基因组密码子为研究对象，运用 CodonW 软件对杜仲的 320 个蛋白编码基因进行同义密码子相对使用频率（RSCU）分析、ENC-GC_{3S}关联分析编码基因的密码子 ENC 值、PR2-plot 偏倚分析编码基因的密码子碱基使用频率。运用 CUSP 软件与 Codon Usage Database 软件对杜仲基因密码子的 GC 含量、使用频率与代表性物种烟草、拟南芥、大肠杆菌和酿酒酵母的密码子 GC 含量和使用频率进行比较。结果表明：杜仲基因密码子的 RSCU>1 的密码子有 30 个，其中 18 个以 G/C 结尾、12 个以 A/U 结尾；说明杜仲基因密码子偏好以 G/C 结尾，且偏好性较强。有效密码子数（ENC)范围为 30～60，该范围内的密码子距离标准曲线较近其 ENC 值小，偏好性较强。PR2-plot 偏倚分析碱基使用频率显示 G>C、U>A。杜仲与代表性物种的 GC 含量分析显示杜仲的 GC_{12}、GC_3 以及平均 GC 含量均高于代表性物种；杜仲与代表性物种的密码子使用频率分析表明杜仲与烟草、酿酒酵母的密码子偏好较为接近，杜仲与拟南芥、大肠杆菌的密码子偏好差距较大。杜仲是我国特有的珍贵中药材对其进行密码子使用模式分析，研究其密码子偏好规律，为杜仲植物基因工程中外源基因的改良及表达提供理论基础。

关键词：杜仲基因，密码子，偏好性，使用模式，代表性物种

中图分类号：Q943　　　　**文献标识码：**A

开放科学(资源服务)标识码(OSID)：



Analysis of codon usage pattern of *Eucommia ulmoides*

WANG Yan[1], ZHAO Yichen[1*], ZHAO Degang[1,2]

(1. *College of Life Sciences,Tea college of Sciences, Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region Ministry of Education, Guizhou University*, Guiyang 550000, China; 2. *Guizhou Academy of Agricultural Sciences*, Guiyang 550000, China)

Abstract: To comprehend the codon usage pattern of genes in *Eucommia ulmoides*, taking the *Eucommia ulmoides* genome codons as the research object, the 320 coding DNA sequences was

基金项目：国家自然科学基金(31660076) ；贵州省教育厅青年人才成长项目(黔教科 KY[2016]126) ；贵州省科技计划项目（黔科合平台人才[2017]5788，[2017]5788-72 号） [Supported by the National Natural Science Foundation of China(31660076); Guizhou Provincial Department of Education Youth Talent Growth Project (KY[2016]126); Guizhou Science and Technology Plan Project ([2017]5788 ,[2017]5788-72)]。

作者简介：王艳（1996-），女，贵州贵阳人，硕士研究生，研究方向为生物化学分子生物学，(E-mail)304014437@qq.com。

通信作者：赵懿琛，博士，副教授，主要从事药用植物（杜仲、花椒、黄姜等）次生产物生物合成及代谢途径调控研究，（E-mail）yczhao@gzu.edu.cn。

analyzed to obtain to the results of relative synonymous codon usage(RSCU), ENC-GC_{3S} analysis of codon ENC values of coding genes and PR2-plot bias analysis of the codon base usage frequency of the coding gene. By using CUSP and Codon Usage Database software to compare the GC contents and codon occurrence frequency of *Eucommia ulmoides* with *Arabidopsis thaliana*, *Nicotiana tabacum*, *Escherichia coli* and *Saccharomyces cerevisiae*. The results showed that the number of RSCU value was great than 1 has 30 codons, including the ending of 18 genes were G/C and 12 genes A/U. These suggested that the codons preferred ends with G/C, and had a strong bias. The number of effective codons (ENC) ranges from 30 to 60. The codons within this range were closer to the standard curve and their ENC values were smaller and had a higher preference. PR2-plot analysis showed that G>C,U>A, in the base usage frequency. Analysis of GC contents of *Eucommia ulmoides* and representative species showed that the *Eucommia ulmoides* GC_{12}, GC_3 and average GC contents were higher than those of representative species. The analysis of the codon usage frequency of *Eucommia ulmoides* and representative species showed that the codon preferences of *Eucommia ulmoides* and *Nicotiana tabacum* and *Saccharomyces cerevisiae* were close.The codon preferences of *Eucommia ulmoides* and *Arabidopsis thaliana* and *Escherichia coli* were quite different. *Eucommia ulmoides* is a precious Chinese medicinal material unique to our country. It analyzes the codon usage pattern and studies its codon preference and provides a theoretical basis for the improvement and expression of foreign genes in *Eucommia ulmoides* plant genetic engineering.

Key word: *Eucommia ulmoides* gene, codon, preference, usage pattern, representative species

蛋白质的基本组成单位是氨基酸，氨基酸由三联体密码子组成。生物体内含有 20 种氨基酸由 61 个密码子编码，剩余 3 个为终止密码子（李滢等，2016）。编码氨基酸的三联体密码子中前两位碱基基本决定氨基酸的种类，第三位碱基的改变对氨基酸影响较小。即在编码氨基酸过程中即便密码子第三位碱基由于基因突变或其他原因造成改变，也不影响氨基酸的正常编码（赵春丽等，2019）。编码同一氨基酸的不同密码子称同义密码子，该现象称为密码子简并性（苏慧等，2015）。密码子简并性具有重要生物学意义，可以减少有害突变；也可使 DNA 上碱基组成有较大变动余地，因此密码子简并性在物种的稳定上起一定作用（张文娟，2006）。理论上在不存在外界压力或者其他干扰因素条件下，同义密码子的使用频率是相同的。但实际在编码氨基酸时同义密码子的使用频率具有差异，这种现象称为密码子偏好性（胡桂兵等，2006）。密码子偏好性在异源表达方面具有重要作用，不仅可以用来描述基因进化规律、基因表达过程中的调控机制，也可用于预测外源基因的最适宿主，通过密码子优化提高宿主中外源基因的表达量以及评估基因的表达水平等（Wang et al., 2013；胡振民等，2019）。生物在长时间的进化过程中由于基因突变和外界环境的影响会形成自身偏好的密码子系统，当某一密码子在编码氨基酸受阻碍时，与其编码相同氨基酸的同义密码子能够代替其行使功能从而保证蛋白质合成的准确性（Sophiarani et al., 2019）。分析生物对同义密码子的使用情况可为该生物的外源基因插入及表达提供一定的理论依据(雷慧等，2019)。

杜仲（*Eucommia ulmoides*）为杜仲科（Eucommiaceae）植物，是我国特有树种，经济价值很高、资源稀少（Liu et al., 2005）。随着基因工程技术的蓬勃发展关于杜仲的研究范围越来越广。近年来对杜仲的研究有杜仲几丁质酶基因在作物真菌病害方面的抗性机制研究（郭林霞等，2016；丁延庆，2017）、杜仲 dirigent 基因的功能分析（王维东等，2018）以及杜仲胶合成相关基因的克隆及分析（王宇等，2015）研究等方面。近些年在杜仲遗传转化方面的研究越来越多，但杜仲为木本植物生长周期长、遗传转化不易进行等原因，因此探究杜仲基因的密码子使用模式可为杜仲的遗传转化提供更多可能性。该研究在实验室前期的

研究基础上对杜仲的 320 个蛋白编码基因进行同义密码子相对使用频率、ENC 与 GC_{3s} 关联分析及 PR2-plot 偏倚分析，并将杜仲与代表性物种大肠杆菌、拟南芥、酿酒酵母和烟草的密码子 GC 含量及密码子使用频率进行比较，分析杜仲与他们对于同义密码子的使用模式。该研究为杜仲基因寻找合适的表达系统提供理论依据，亦可为外源基因在杜仲中的异源表达提供理论基础。

1 材料与方法

1.1 实验材料

以杜仲基因组测序所得序列为研究对象。杜仲基因组序列为实验室前期基础通过试剂盒法提取杜仲总 RNA，经反转录得 cDNA，测序获得 320 个蛋白编码基因。大肠杆菌、拟南芥、烟草、酿酒酵母的密码子使用频率数据来源于 Codon Usage Database 软件。

1.2 实验方法

1.2.1 同义密码子相对使用频率分析

同义密码子相对使用频率（relative synonymous codon usage, RSCU）指在编码氨基酸时指定密码子与所有能编码该氨基酸的密码子的比值。RSCU值表明密码子使用偏离随机选择的程度，反映同义密码子非均衡使用的偏好程度。RSCU<1 说明编码氨基酸时使用较多随机密码子，偏好性较弱；RSCU=1，表明编码氨基酸的指定密码子数等于所有能编码该氨基酸的密码子数，即该密码子不具偏好性；若RSCU>1 说明编码氨基酸时使用较多的指定密码子，其偏好性较强。该分析用CodonW(http://bioweb.pasteur. Fr/seqanal/interfaces/codonw.html)软件分析基因序列的同义密码子相对使用次数和频率，来评估密码子的偏好性（Subramanian, 2008）。

1.2.2 ENC-GC_{3s} 关联分析

有效密码子数（effective number of codons, ENC）指在编码氨基酸时所用的全部密码子的种类个数，其最小值为 20，最大值为 61。ENC 值越接近 20 说明密码子偏好性越强，越靠近 61 则偏好性越弱。以 ENC 值为纵坐标，GC_3 值为横坐标作图分析。图中标准曲线代表无选择压力时，密码子偏好性完全由突变所决定（Pamela et al., 2007）。散点分布于曲线周围，散点与标准曲线的距离表明密码子的偏好性强弱。距离远表示 ENC 值小密码子偏好性强；反之则 ENC 值大密码子偏好性弱。

1.2.3 PR2-plot 偏倚分析

通常在无选择压力情况下，密码子 A/U、G/C 含量分布均匀（张雅丽等，2002）。但因外界压力、环境等因素的影响它们的含量分布不均，PR2-plot 图被分为 4 个区域，在平面图的上方表示碱基 A 的使用频率多于碱基 U，反之 U 多于 A；在平面图的左方表示碱基 C 的使用频率高于碱基 G，反之 G 多于 C。

1.2.4 杜仲与代表性物种的 GC 含量及密码子使用频率比较

利用分析软件 CUSP 计算杜仲基因中各密码子的出现频率，并与从 Codon Usage Database 软件中获得的烟草、拟南芥、大肠杆菌、酿酒酵母的密码子使用频率进行比较，比值≥2 或≤0.5 表示两物种间密码子使用偏好性差异较大（晁岳恩等，2012）。

2 结果与分析

2.1 RSCU 值分析

通常以 RSCU 值来分析密码子偏好性，该研究运用 CodonW 软件分析杜仲基因密码子的 RSCU 值，RSCU 值指编码氨基酸时指定密码子与所有能编码该氨基酸的密码子的比值。分析杜仲基因密码子的 RSCU 值结果得，RSCU>1 的密码子有 30 个，其中以 UGA、CUC、AGG 偏好性最强，RSCU 值分别为 1.61、1.51 和 1.46。RSCU>1 的密码子有 18 个以 G/C 结尾，12 个以 A/U 结尾，说明杜仲基因密码子偏爱以 G/C 结尾。RSCU=1 的密码子有 4 个，RSCU<1 的密码子有 29 个，其中包括 2 个终止密码子；从 RSCU 值分析得出杜仲基因在编码氨基酸时对同义密码子的使用偏好性较强。

表 1 杜仲基因编码氨基酸的同义密码子相对使用度

Table 1 Relative usage of synonymous codons encoding amino acids in *Eucommia ulmoides*

氨基酸 Amino acid	密码子 Codon	个数 Count	相对使用度（RSCU）	氨基酸 Amino Acid	密码子 Codon	个数 Count	相对使用度（RSCU）
苯丙氨 Phe	UUU	857	0.90	酪氨酸 Tyr	UAU	308	1.01
	UUC	1 057	1.10		UAC	299	0.99
亮氨酸 Leu	UUA	279	0.38	终止密码 TER	UAA	67	0.63
	UUG	968	1.32		UAG	81	0.76
	CUU	948	1.29		UGA	172	1.61
	CUC	1 108	1.51	组氨酸 His	CAU	610	0.93
	CUA	312	0.43		CAC	695	1.07
	CUG	789	1.07	谷氨酰胺 Gln	CAA	620	1.04
异亮氨 Ile	AUU	687	1.00*		CAG	570	0.96
	AUC	905	1.32	天冬酰胺 Asn	AAU	657	0.98
	AUA	470	0.68		AAC	684	1.02
甲硫氨 Met	AUG	1 216	1.00*	赖氨酸 Lys	AAA	728	0.93
缬氨酸 Val	GUU	693	1.10		AAG	835	1.07
	GUC	601	0.96	天冬氨酸 Asp	GAU	734	1.21
	GUA	455	0.72		GAC	479	0.79
	GUG	763	1.21	谷氨酸 Glu	GAA	760	0.92
丝氨酸 Ser	UCU	824	1.11		GAG	900	1.08
	UCC	1 035	1.39	半胱氨酸 Cys	UGU	400	0.91
	UCA	693	0.93		UGC	482	1.09
	UCG	730	0.98	色氨酸 Trp	UGG	703	1.00*
	AGU	434	0.58	精氨酸 Arg	CGU	288	0.54
	AGC	756	1.01		CGC	428	0.80
脯氨酸 Pro	CCU	613	0.86		CGA	439	0.82

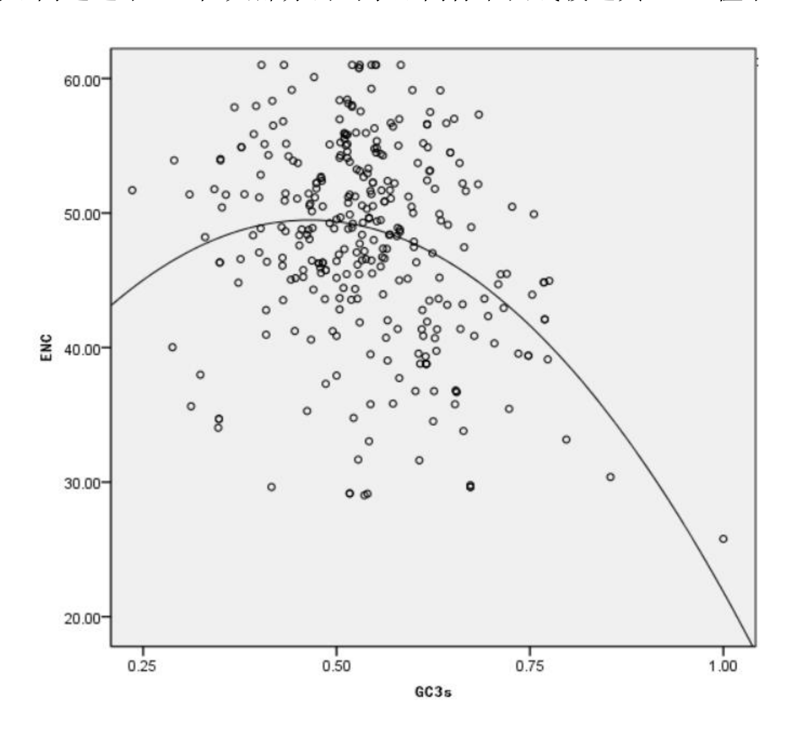
	CCC	638	0.89		CGG	580	1.09
	CCA	811	1.13		AGA	687	1.29
	CCG	805	1.12		AGG	776	1.46
苏氨酸 Thr	ACU	498	0.85	甘氨酸 Gly	GGU	588	0.95
	ACC	701	1.19		GGC	616	1.00*
	ACA	603	1.03		GGA	663	1.07
	ACG	548	0.93		GGG	603	0.98
丙氨酸 Ala	GCU	692	1.03				
	GCC	800	1.20				
	GCA	621	0.93				
	GCG	564	0.84				

注：表中有下划线的为 RSCU>1 的值，标*的为 RSCU=1 的值。

Note: The underlined values in the table are RSCU>1, * is the value of RSCU=1.

2.2 ENC 与 GC_{3s} 关联分析

以杜仲基因密码子 ENC 值为纵坐标，GC_3 值为横坐标作图分析，如图 1。ENC 值表示密码子的偏好性强弱，ENC 值越接近 20 说明密码子偏好性越强；ENC 值越接近 61 说明密码子偏好性越弱。分析得杜仲基因 ENC 值多集中分布于 30～60 之间；该区域内的密码子距离标准曲线的距离远近不一，但大部分密码子距离标准曲线较近其 ENC 值小，偏好性较强。



横坐标为 GC_3 值，纵坐标为 ENC 值。

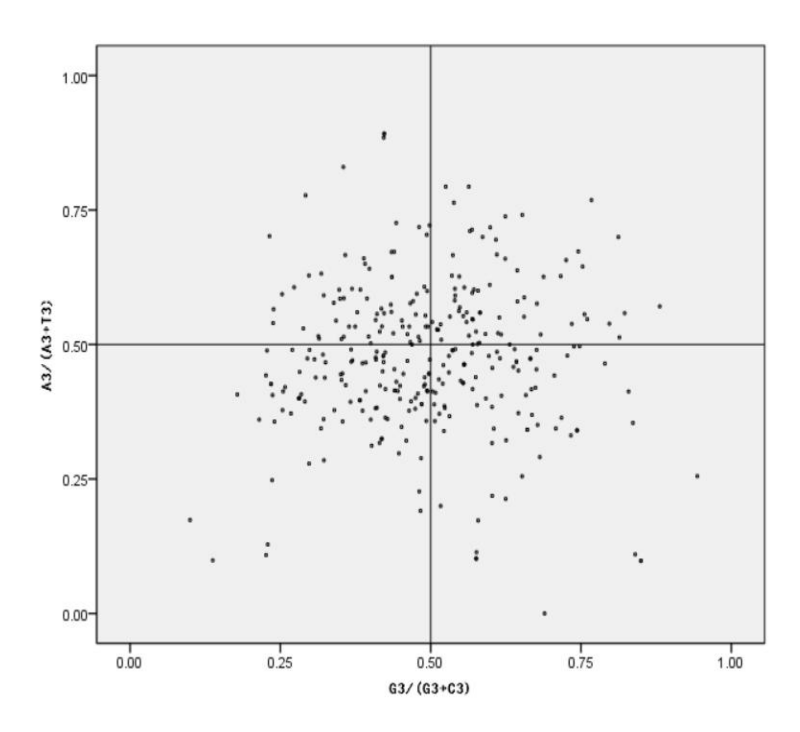
The abscissa is the GC_3 value and the ordinate is the ENC value.

图 1 ENC-plot 分析

Fig.1 ENC-plot analysis

2.3 PR2-plot 偏倚分析

对杜仲基因密码子进行 PR2-plot 偏倚分析，如图 2 所示。图 2 中密码子多分布于平面图的下方和平面图的右方，说明碱基使用频率 G>C、U>A。若密码子偏好性完全受突变的影响，则碱基 A、U、C、G 的使用频率相当，因此说明杜仲基因密码子受到突变和选择的影响。



横坐标为 $G_3/(G_3+C_3)$，纵坐标为 $A_3/(A_3+T_3)$。

The abscissa is $G_3/(G_3+C_3)$ and the ordinate is $A_3/(A_3+T_3)$.

图 2 PR2-plot 偏倚图分析

Fig. 2 Analysis of PR2 bias plot

2.4 杜仲与代表性物种的 GC 含量及密码子使用频率比较

将杜仲与 4 种代表性物种的 GC 含量进行比较，如表 2 所示。杜仲基因密码子的平均 GC 含量最高为 51.7%，密码子第三位碱基 GC 含量也最高为 53.61%，说明杜仲基因密码子偏好 GC 结尾。其次为大肠杆菌、拟南芥、烟草和酿酒酵母。杜仲与代表性物种的密码子使用频率以及频率比值结果如表 3 所示，杜仲与代表性物种的密码子使用频率比值≥2 或≤0.5 的分别有 31 个、34 个、34 个、32 个，该值说明两物种间对该密码子的使用偏差，表 3 中杜仲与大肠杆菌和拟南芥的密码子使用偏好具有明显差异，而杜仲与烟草、酿酒酵母的密码子偏好性差别较小。若要为杜仲基因寻找宿主表达系统的话，可选择烟草或酿酒酵母较为合适。

表 2 杜仲与代表性物种 GC 的含量

Table 2 Contents of GC in *Eucommia ulmoides* and representative species

	杜仲 *Eucommia ulmoides*	拟南芥 *Arabidopsis thaliana*	烟草 *Nicotiana tabacum*	大肠杆菌 *Escherichia coli*	酿酒酵母 *Saccharomyces cerevisiae*

GC_{12}	50.85%	45.69%	45.3%	47.22%	40.61%
GC_3	53.61%	42.38%	39.53%	47.45%	38.1%
平均含量 Average content	51.70%	44.59%	43.38%	47.3%	39.77%

表 3 杜仲与代表性物种的密码子使用频率比较

Table 3 Comparison of codon frequency between *Eucommia ulmoides* and representative species

氨基酸 Amino acid	密码子 Codon	密码子使用频率 Frequency of codon usage (%)					频率比值 Comparison of codon preference			
		Eu	Nt	At	Eco	Sc	Eu/Nt	Eu/At	Eu/Eco	Eu/Sc
苯丙氨酸 Phe	UUU	9.9	25.1	21.8	24.4	26.1	0.39	0.45	0.4	0.37
	UUC	20.9	18.0	20.7	13.9	18.4	1.16	1.00	1.50	1.13
亮氨酸 Leu	UUA	3.8	13.4	12.7	17.4	26.2	0.28	0.29	0.21	0.14
	UUG	33.6	22.3	20.9	12.9	27.2	1.50	1.60	2.60	1.23
	CUU	4.9	24.0	24.1	14.5	12.3	0.20	0.20	0.33	0.39
	CUC	4.9	12.3	16.1	9.5	5.4	0.39	0.30	0.51	0.90
	CUA	2.2	9.4	9.9	5.6	13.4	0.23	0.22	0.39	0.16
	CUG	13.2	10.2	9.8	37.4	10.5	1.29	1.34	0.35	1.28
异亮氨酸 Ile	AUU	10.4	27.8	21.5	29.6	30.1	0.37	0.48	0.35	0.34
	AUC	11	13.9	18.5	19.4	17.2	0.79	0.59	0.56	0.63
	AUA	3.8	14.0	12.6	13.3	17.8	0.27	0.30	0.28	0.21
甲硫氨酸 Met	AUG	38.5	25.0	24.5	23.7	20.9	1.54	1.57	1.62	1.84
缬氨酸 Val	GUU	6.0	26.8	27.2	21.6	22.1	0.22	0.22	0.27	0.27
	GUC	4.4	11.1	12.8	13.1	11.8	0.39	0.34	0.33	0.37
	GUA	2.7	11.4	9.9	13.1	11.8	0.23	0.27	0.20	0.22
	GUG	18.1	16.7	17.4	19.9	10.8	1.08	1.04	0.90	1.67
丝氨酸 Ser	UCU	25.8	20.0	25.2	13.1	23.5	1.29	1.02	1.96	1.09
	UCC	29.7	10.2	11.2	9.7	14.2	2.91	2.65	3.06	2.09
	UCA	44	17.6	18.3	13.1	18.7	2.5	2.40	3.35	2.35
	UCG	53.4	5.3	9.3	8.2	8.6	10.07	5.74	6.51	6.20
	AGU	7.7	13.3	14	13.2	14.2	0.57	0.55	0.58	0.54
	AGC	13.7	10.0	11.3	14.3	9.8	1.37	1.21	0.95	1.39
脯氨酸 Pro	CCU	11.5	18.7	18.7	9.5	13.5	0.61	0.61	1.21	0.85
	CCC	19.2	6.6	5.3	6.2	6.8	2.90	3.62	3.09	2.82
	CCA	36.9	19.8	16.1	9.1	18.3	1.86	2.29	4.05	2.01
	CCG	62.2	5.0	8.6	14.5	5.3	12.44	7.23	4.28	11.73
苏氨酸 Thr	ACU	12.6	20.3	17.5	13.1	20.3	0.62	0.72	0.96	0.62
	ACC	25.3	9.7	10.3	18.9	12.7	2.60	2.45	1.33	1.99

	ACA	22	17.4	15.7	15.1	17.8	1.26	1.40	1.45	1.23
	ACG	30.3	4.5	7.7	13.6	8	6.73	3.93	2.22	3.78
丙氨酸 Ala	GCU	11	31.2	28.3	18.9	21.2	0.35	0.38	0.58	0.51
	GCC	20.9	12.5	10.3	21.6	12.6	1.67	2.02	0.96	1.65
	GCA	19.8	23.1	17.5	23	16.2	0.85	1.13	0.86	1.22
	GCG	45.7	5.8	9	21.1	6.2	7.87	5.07	2.16	7.37
酪氨酸 Tyr	UAU	3.8	17.8	14.6	21.6	18.8	0.21	0.26	0.17	0.20
	UAC	8.2	13.5	13.7	11.7	14.8	0.60	0.59	0.70	0.55
组氨酸 His	CAU	6.0	13.4	13.8	12.4	13.6	0.44	0.43	0.48	0.44
	CAC	6.0	8.7	8.7	7.3	7.8	0.68	0.68	0.82	0.76
谷氨酰胺 Gln	CAA	6.0	20.7	19.4	14.4	27.3	0.28	0.30	0.41	0.21
	CAG	12.6	15.0	15.2	26.7	12.1	0.84	0.82	0.47	1.04
天冬酰胺 Asn	AAU	7.1	28.0	22.3	29.3	35.7	0.25	0.31	0.24	0.19
	AAC	12.6	17.9	20.9	20.3	24.8	0.70	0.60	0.62	0.50
赖氨酸 Lys	AAA	11	32.6	30.8	37.2	41.9	0.33	0.35	0.29	0.26
	AAG	21.4	33.5	32.7	15.3	30.8	0.63	0.65	1.39	0.69
天冬氨酸 Asp	GAU	6.6	36.9	36.6	33.7	37.6	0.17	0.18	0.19	0.17
	GAC	6.0	16.9	17.2	17.9	20.2	0.35	0.34	0.33	0.29
谷氨酸 Glu	GAA	11.5	36	34.3	35.1	45.6	0.31	0.33	0.32	0.25
	GAG	22	29.4	32.2	19.4	19.2	0.74	0.68	1.13	1.14
半胱氨酸 Cys	UGU	10.4	9.8	10.5	5.9	8.1	1.06	0.99	1.76	1.28
	UGC	6.6	7.2	7.2	5.5	4.8	0.91	0.91	1.2	1.37
色氨酸 Trp	UGG	24.2	12.2	12.5	13.4	10.4	1.98	1.93	1.80	2.32
精氨酸 Arg	CGU	4.4	7.5	9	15.9	6.4	0.58	0.48	0.27	0.68
	CGC	3.3	3.9	3.8	14	2.6	0.84	0.86	0.23	1.26
	CGA	9.9	5.3	6.3	4.8	3	1.86	1.57	2.06	3.3
	CGG	11.0	3.7	4.9	7.9	1.7	2.97	2.24	1.39	6.47
	AGA	17.6	12.2	19	7.1	21.3	1.44	0.92	2.47	0.82
	AGG	38	7.5	11	4	9.2	5.06	3.45	9.50	4.13
甘氨酸 Gly	GGU	3.8	22.3	22.2	23.7	23.9	0.17	0.17	0.16	0.15
	GGC	7.1	11.2	9.2	12.3	9.8	0.63	0.77	0.57	0.72
	GGA	11	23.2	24.2	13.6	10.9	0.47	0.45	0.80	1.00
	GGG	20.3	10.5	10.2	20.6	6	1.93	1.99	0.98	3.38

注：表中 Eu、Nt、At、Eco、Sc 分别代表杜仲、烟草、拟南芥、大肠杆菌和酿酒酵母；表中有横线的为两物种间密码子使用频率大于等于 2 或小于等于 0.5 的值。

Note: Eu, Nt, At, Eco and Sc in the table respectively represent *Eucommia ulmoides*, *Nicotiana tabacum*, *Arabidopsis thaliana*, *Escherichia coli*, *Saccharomyces cerevisiae*; The vaules with horizontal lines in the table are the values with codon occurrence frequency greater than or equal to 2 or less than or equal to 0.5 between the two species.

3 讨论

氨基酸由三联体密码子编码，密码子简并性可使碱基第三位改变的情况下编码正确的氨

基酸，保证蛋白质合成的准确。不同生物的密码子使用偏好存在差异，即便同一生物不同基因的密码子使用偏好也存在差异（吴宪明等，2007）。密码子使用偏好是生物在进化过程中突变、选择的综合结果，研究表明密码子偏好与基因表达水平相关，高表达基因倾向使用最优密码子（赵洋等，2011）。密码子偏好性也反应了生物间进化的关系，可作为生物进化关系的参考（陈哲等，2017）。密码子使用模式分析可为生物的外源基因插入及表达提供一定的理论依据（罗洪等，2015）。杜仲是我国特有的珍贵药材，具有降血压、安胎等功效。但因其为木本植物、生长周期长等原因，杜仲遗传转化不易进行且转基因杜仲移栽至土里存活率不高。因此探究杜仲基因的密码子使用模式可为杜仲的遗传转化提供更多可能性。

该研究以杜仲基因组为对象，分析杜仲基因密码子使用模式。结果表明杜仲基因密码子RSCU>1 的有 30 个，其中 18 个密码子以 G/C 结尾、12 个密码子以 A/U 结尾；说明杜仲基因密码子偏好以 G/C 结尾，这种偏好可能是由于以 G/C 结尾的密码子具有较高的结合能，有利于保证翻译的准确性（蒋玮等，2014)。ENC 值大小可量化一个基因的密码子使用频率与同义密码子平均使用频率的偏差，从而检测单个基因密码子的偏好程度(朱婷婷等，2017)。ENC 值与期望 ENC 值差异越小，说明突变对密码子偏好性影响越大；ENC 值与期望 ENC 值差异越大，说明选择对密码子偏好性影响越大（杨国锋等，2015）。分析杜仲基因密码子 ENC 值，其值集中分布于 30-60 间，且该区域内的密码子距离标准曲线较近，偏好性较强。PR2-plot 分析中，平面图显示各基因的碱基组成，其中点代表无偏倚性使用时的密码子状态，即 A=U，G=C（尚明照等，2011；邢朝斌等，2013）。杜仲基因密码子的 PR2-plot 分析显示碱基使用频率 G>C、U>A。说明杜仲基因密码子的使用偏好受到自然选择和基因突变的影响。杜仲与拟南芥、烟草、大肠杆菌、酿酒酵母等代表性物种的密码子使用频率都具有不同程度的差异，若这些生物的基因要在杜仲中表达，则应先进行密码子优化。杜仲与烟草、酿酒酵母的密码子偏好性差别较小，与大肠杆菌、拟南芥的密码子偏好性差别较大。如果烟草、酿酒酵母的基因要在杜仲中表达则只需优化个别密码子，若要为杜仲基因寻找宿主系统则烟草、酿酒酵母较为合适。

随着基因工程的广泛应用，外源基因的功能验证多采用转基因异源表达。外源基因导入宿主系统后由于其密码子与宿主系统的密码子使用偏好不同，导致表达效率低下。密码子偏好性分析在基因异源表达方面具有重要意义，基于同义密码子的使用频率不同为了提高外源基因的表达水平可对稀有密码子进行改造，增加优势密码子使用频率使得外源基因更好的表达(胡晓艳等，2019)。该研究分析珍贵中药材杜仲的基因密码子使用模式，对于杜仲中外源基因的插入及高效表达具有意义并提供了理论基础。

参 考 文 献：

CHAO YE, CHANG Y, WANG MF, 2012. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species[J]. Acta Agric Boreal-Sin, 27(4): 60-64. [晁岳恩，常阳，王美芳，2012. 7 种作物叶绿体基因的密码子偏好性及聚类分析[J]. 华北农学报，27(4)：60-64.]

CHEN Z, HU FC, WANG XH, et al., 2017. Analysis of codon bias of *Ananas comosus* with genome sequencing data[J]. J Fruit Sci, 34(8): 946-955. [陈哲，胡福初，王祥和，等，2017. 菠萝密码子使用偏好性分析[J]. 果树学报，34(8)：946-955.]

DING YQ, 2017. Transformed *EuCHIT1* gene in wheat (*Triticum aestivum*) enhances resistance to stripe rust and leaf rust[D]. Guiyang: Guizhou University: 21. [丁延庆，2017. 利用杜仲几丁质酶基因 *EuCHIT1* 创制抗锈病小麦新种质的研究[D]. 贵阳：贵州大学：21.]

GUO LX, DONG X, ZHAO DG, 2016. Transgenic tomato plants expressing a *Eucommia ulmoides* chitinase gene *EuCHIT1* and their resistance to *Botrytis cinerea* [J]. Plant Physiol J, 52(5): 703-714. [郭林霞，董旋，赵德刚，

2016. 转杜仲几丁质酶基因 *EuCHIT1* 番茄提高对灰霉病的抗性研究[J]. 植物生理学报，52(5)：703-714.]

HU GB, ZHANG SL, XU CJ, 2006. Analysis of codon usage between different citrus species[J]. J S Chin Agric Univ, 27(1): 2-3. [胡桂兵，张上隆，徐昌杰，2006. 不同种类柑橘的密码子用法分析[J]. 华南农业大学学报，27(1)：2-3.]

HU XY, XU YQ, HAN YZ, et al., 2019. Codon usage bias analysis of the chloroplast genome of *Ziziphus jujuba* var. *spinosa*[J]. J Forest Environ, 39(6): 621-628. [胡晓艳，许艳秋，韩有志，等，2019. 酸枣叶绿体基因组密码子使用偏性分析[J]. 森林与环境学报，39(6)：621-628.]

HU ZM, WAN Q, LI H, et al., 2019. Analysis of codon usage features of *CsNRT1.1* gene in *Camellia sinensis*[J]. Jiangsu J Agric Sci, 35(4): 896-903. [胡振民，万青，李欢，等，2019. 茶树 *CsNRT1.1* 基因密码子使用特性分析[J]. 江苏农业学报，35(4)：896-903.]

JIANG W, LÜ BB, HE JH, et al., 2014. Codon usage bias in the straw mushroom *Volvariella volvacea*[J]. Chin J Biotechnol, 30(9): 1424-1435. [蒋玮，吕贝贝，何建华，等，2014. 草菇密码子偏好性分析[J]. 生物工程学报，30(9)：1424-1435.]

LEI H, LI G, WANG NY, 2019. Analysis of codon usage bias in the chloroplast genome of *Trollius chinensis* bunge[J]. J Shanxi Agric Sci, 47(08): 1300-1305. [雷慧，李鸽，王娜玉，2019.金莲花叶绿体基因组密码子偏好性分析[J]. 山西农业科学，47(08)：1300-1305.]

LI Y, KUANG XJ, SUN C, 2016. Analysis of codon usage bias based on *Fritillaria cirrhosa* transcriptome[J]. Chin J Chin Mat Med, 41(11): 3-5. [李滢，匡雪君，孙超，2016. 川贝母转录组密码子使用偏好性分析[J]. 中国中药杂志，41(11)：3-5.]

LIU Q, DOU S, JI Z, et al., 2005. Synonymous codon usage and gene function are strongly related in *Oryza sativa*[J]. Biosystems, 80(2): 123-131.

LUO H, HU SS, WU Q, et al., 2015. Analysis of buckwheat chloroplast gene codon bias[J]. Genomic Appl Biol, 34(11): 2457- 2464. [罗洪，胡莎莎，吴琦，等，2015. 甜荞叶绿体基因密码子偏爱性分析[J]. 基因组学与应用生物学，34(11)：2457- 2464.]

PAMELA M, SURAJIT B, TAPASH C G, 2007. Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes[J]. Gene, 400(1): 3-13.

SHANG MZ, LIU F, HUA JP, et al., 2011. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*[J]. Sci Agric Sin, 44(2): 245-253. [尚明照，刘方，华金平，等，2011. 陆地棉叶绿体基因组密码子使用偏性的分析[J]. 中国农业科学，44(2)：245-253.]

SOPHIARANI Y, ARIF U, SUPRIYO C, 2019. Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. *assamica* and *Camellia sinensis* var. *sinensis* in comparison to *Camellia pubicosta*[J]. J Integr Agric, 18(12): 2775-2780.

SUBRAMANIAN S, 2008. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes[J]. Genetics, 178(4): 2429-2432.

SU H, LI YG, TAN WY, 2015. Analysis of codon usage bias and optimization of codon of *gdh A* gene in *Escherichia coli*[J]. Genomic Appl Biol, 34(3): 521-529. [苏惠，李永光，谭文雍，2015. 大肠杆菌 *gdhA* 基因的密码子偏好性分析及优化[J]. 基因组学与应用生物学，34(3)：521-529.]

WANG WD, ZHAO DG, ZHAO YC, 2018. Cloning and sequence analysis of dirigent protein coding genes from *Eucommia ulmoides*[J]. Genomic Appl Biol, 37(3): 1293-1301. [王维东，赵德刚，赵懿琛，2018. 杜仲 Dirigent 蛋白编码基因的克隆及序列分析[J]. 基因组学与应用生物学，37(3)：1293-1301.]

WANG X, WU L, ZHOU P, et al., 2013. Analysis of synonymous codon usage patterns in the genus *Rhizobium*[J]. World J Microbiol Biotechnol, 29(11): 2015-2024.

WANG Y, DONG X, ZHAO DG, 2015. Molecular cloning and analysis of *EuREF1* involved into rubber biosynthesis in *Eucommia ulmoides* Olive[J]. Genomic Appl Biol, 34(5):917-925. [王宇，董旋，赵德刚，

2015. 杜仲胶合成相关基因 *EuREF1* 的克隆及分析[J]. 基因组学与应用生物学，34(5)：917-925.]

WU XM, WU SF, REN DM, et al., 2007. The analysis method and progress in the study of codon bias[J]. Hereditas, 29(4): 420-426. [吴宪明，吴松锋，任大明，等，2007. 密码子偏性的分析方法及相关研究进展[J]. 遗传，29(4)：420-426.]

XING ZB, CAO L, ZHOU M, et al., 2013. Analysis on codon usage of chloroplast genome of *Eleutherococcus senticosus*[J]. Chin J Chin Mat Med, 38(5): 661-665. [邢朝斌，曹蕾，周秘，等，2013. 刺五加叶绿体基因组密码子的用法分析[J]. 中国中药杂志，38(5)：661-665.]

YANG GF, SU KL, ZHAO YR, et al., 2015. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*[J]. Acta Pratac Sin, 24(12): 171-179. [杨国锋，苏昆龙，赵怡然，等，2015. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. 草业学报，24(12)：171-179.]

ZHANG WJ, 2006. Codon analysis and its application in bioinformaties and evolutionarystudies[D]. Shanghai: Fudan University: 20. [张文娟，2006. 基于密码子水平的生物信息学分析及进化研究[D]. 上海：复旦大学：20.]

ZHAO CL, PENG LY, WANG X, et al., 2019. Codon bias and evolution analysis of *AtGAI* in *Amaranthus tricolor* L.[J]. J Chin Agric Univ(Soc Sci Ed), 24(12): 10-22. [赵春丽，彭丽云，王晓，等，2019. 苋菜 *AtGAI* 基因密码子偏好性与进化分析[J]. 中国农业大学学报，24(12)：10-22.]

ZHANG YL, YANG GQ, GUO Y, 2002. Expression of bovine protamine gene optimized by codon in *E. coli*[J]. Chin High Technol Lett, 12(4): 42-46. [张雅丽，杨国庆，郭英，2002. 密码子优化的牛精蛋白基因在大肠杆菌中的表达[J]. 高技术通讯，12(4)：42-46.]

ZHAO Y, YANG Y, LIU Z, et al., 2011. Analysis of codonusage in tea plant (*Camellia sinensis*)[J]. J Tea Sci, 31(4): 319-325. [赵洋，杨阳，刘振，等，2011. 茶树密码子用法分析[J]. 茶叶科学，31(4)：319-325.]

ZHU TT, ZHANG L, CHEN WS, et al., 2017. Analysis of chloroplast genomes in 1 342 plants[J]. Genomic Appl Biol, 36(10): 4323- 4333. [朱婷婷，张磊，陈万生，等，2017. 1 342 个植物叶绿体基因组分析[J]. 基因组学与应用生物学，36(10)：4323- 4333.]